International Journal of Artificial Intelligence and Applications (IJAIA), Vol.9, No.4, July 2018

PERFORMANCE EVALUATION OF VARIOUS EMOTION CLASSIFICATION APPROACHES FROM PHYSIOLOGICAL SIGNALS

Dilana Hazer-Rau, Lin Zhang, Harald C. Traue

Medical Psychology, University of Ulm, Ulm, Germany
dilana.hazer@uni-ulm.de

ABSTRACT

This paper aims at evaluating the performance of various emotion classification approaches from psychophysiological signals. The goal is to identify the combinations of approaches that are most relevant for assessing human affective states. A classification analysis of various combinations of feature selection techniques, classification algorithms and evaluation methods is presented. The emotion recognition is conducted based on four physiological signals: two electromyograms, skin conductivity and respiration sensors. Affective states are classified into three different emotion classes: 2-category-class (Arousal), 3-category-class (Valence) and 5-category-class (Valence/Arousal). The performance of the various combinations of approaches is evaluated by comparing the resulting recognition rates. For all the category-classes, the best results are obtained when considering skin conductivity combined with the respiration signals. Highest rates when fusing all physiological channels resulted when applying the SFS feature selection, the LDA classifier and the normal split evaluation approach, showing a robust combination of approaches leading to good performance.

KEYWORDS

Affective Computing, Emotion Recognition, Feature Extraction & Selection, Classification Algorithms, Biosignal Analysis.

1. INTRODUCTION

Assessing individual human's emotions can be useful in many scientific areas. Amongst these areas are the healthcare and educational fields, mobile and driving applications and the development of cognitive intelligent systems such as companion, prevention or elderly support systems. In order to approach such intelligent applications, one of the important prerequisites is to develop a reliable emotion recognition system, which can guarantee acceptable computational accuracy and adaptability to practical applications.

Human affective states can be assessed based on the analysis of various modalities. Several studies on emotion recognition including facial expression, speech, body gestures, contexts and physiological signals have been performed in the past few decades [1,2,3]. Among these different modalities, psychophysiological signals have various considerable advantages in assessing human affective expressions. For instance, as honest signals, they are considered as the most reliable for human emotion recognition as they cannot be easily triggered by any conscious or intentional control [4].

Various feature extraction and selection optimization approaches, classification algorithms, and evaluation methods are currently been used for the emotion recognition from psychophysiological data in recent years [5,6,7]. To adapt to the fast-changing technologies and recent applications, automatic recognition algorithms have been also applied to different psychophysiological signals in order to efficiently compute and classify human's affective states [8,9]. Thereby, a variety of emotional models have been employed to describe the emotional space for the affect recognition based on discrete and dimensional models. Subsequently, emotion classification based on psychophysiological signals, allows suitable categorization of the affective states, for instance in terms of valence, arousal and dominance (low / high) subspaces [10,11].

Methods to induce various emotional states have been also developed and applied in different ways, including standardized images as induction stimuli, film-based induction to evoke affective and emotional changes, sounds, voices and music stimuli, posed depictions of emotion using facial expressions, bodily movements and posture, revitalization of experienced emotional situations (e.g. imagining the future, remembering the past or creating fictitious imaginings), physiological manipulations (e.g. through pharmacological means) or affective connotations in terms of evocative words [12,13,14,15]. Amongst all the emotion induction methods, visual induction stimuli can be easily categorized in terms of emotional content and are well controlled with regard to the size and duration of the material. Being also easy to integrate within experimental setups, standardized picture stimulus material is therefore often used for the elicitation of various human affective-states.

In the following, we present a classification analysis for the human emotion recognition from psychophysiological signals. The aim of the study is to evaluate the performance of various combinations of emotion classification approaches in order to understand and identify the approaches that are most reliable for identifying different emotional states. The emotion induction is based on standardized image stimuli. The application of various combinations of feature selection techniques, classification algorithms and evaluation methods is thereby investigated and the results are evaluated and compared in terms of classification rates.

2. METHODS

2.1. Subjects Description

The dataset used in this study is based on subjects previously recruited via bulletins distributed on the campus of the University of Ulm. The total sample size was n= 107 subjects (74 women, 33 men) between the age of 20 and 75 years old. Seven subjects had to be excluded from the study due to technical problems and missing data, so that the final number of subjects left and considered in this study is n= 100 subjects. All subjects were right-handed, healthy and had normal vision or corrected normal vision.

2.2. Emotion Elicitation

Emotion induction was conducted using standardized stimuli from the International Affective Picture System (IAPS) and extended by the Ulm Picture Set [16] to represent the VAD (Valence, Arousal, Dominance) space [17]. Both picture systems allow a dimensional induction of emotions according to their ratings in the valence, arousal and dominance dimensions [17]. The experimental design is thereby based on a previous experiment [6], adapted to stimulate prolonged emotion induction. Prolonged presentations consisting of 10 pictures with similar rating à 2s each (total of 20s per presentation) are used to intensify the elicitation [18]. A total of 10 sets of these picture-presentations à 20s each were presented to induce a total of 10 different VAD-states. Thus, the induced VAD-space for the 10 sets of picture-presentations included combinations of positive/negative/neutral (+/-/0) Valence (V), positive/negative (+/-) Arousal (A), and positive/negative (+/-) Dominance (D) values. In order to neutralize the user's affective state between 2 different sets of presentations, 20s of neutral fixation crosses were introduced as baseline inbetween. In total, 100 pictures were used for the emotion induction. While the order of pictures in each presentation-set was fixed, the display of the 10 sets was randomized.

In terms of classification, picture-presentation with similar ratings in terms of Arousal (+/-) and/or Valence (+/-/0) and/or Dominance were combined into one category. In this study, we defined and evaluated three different category-classes presented in Table 1.

Table 1. Overview of the three different category classes used for the evaluation.

2-category-class	Arousal A: + / -
3-category-class	Valence V: + / - / 0
5-category-class	VA: 0- / ++ / -+ / +- / --

2.3. Physiological Signals

The physiological signals analyzed in this study were acquired via electrodes connected to the subjects. They include:

Skin Conductivity (SC):
Two electrodes connected to the sensor were positioned on the index and ring fingers. Since the sweat glands are innervated sympathetically, electrodermal activity is a good indicator of the inner tension of a person.

Respiration (RSP):
The respiration sensor was used to measure the abdominal breathing frequency, as well as the relative depth of breathing. It was placed tight enough in the abdominal area just above the navel.

Electromyography (EMG):
Electrical muscle activity is related to the activity of the sympathetic nervous system. We used two-channel electromyography signals for the zygomaticus major (Zyg.) and the corrugator supercilii, (Corr.) muscles, which are expected to be active during different emotions.

2.4. Data Processing

The processing of the physiological biosignals includes the pre-processing of the raw data, the feature extraction and the emotion classification. These steps are described in the following subsections.

2.4.1. Pre-Pocessing

The raw data were first pre-processed by extracting the relevant signals and picture-sessions from the whole dataset. Then, the extracted data were further processed and prepared to meet the AuBT (Augsburg Biosignal Toolbox) file format requirements [19]. The toolbox provides tools to analyze physiological signals for the emotion recognition [20]. It is used in this study for the later signal processing and analysis including the feature extraction and feature selection as well as the emotion classification and the evaluation analysis. For the application of these pre-processing steps and for the optimization of the quality of the signals, various automation scripts and filtering techniques were additionally composed and implemented as Matlab-based functions.

2.4.2. Feature Extraction and Emotion Classification

In the next step, the AuBT toolbox was used to extract features from the physiological signals including skin conductivity and respiration, as well as the electromyography signals, EMG corrugator and EMG zygomaticus. All acquired signals were thereby examined both individually as well as in various combinations among each other. For each of the resulting signal configuration, the selection of the features was optimized and the resulting selected features were used to train and evaluate a classifier. Feature selection was thereby optimized using the Sequential Forward Selection (SFS) and the Sequential Backward Selection (SBS) algorithms. As classification methods, the k-Nearest-Neighbors (kNN) and the Linear Discriminant Analysis (LDA) models were adopted. Finally, the classifiers were evaluated using three different evaluation methods including the normal split, the random split and the one-leave-out methods.

Each of the mentioned approaches was executed using various combinations of strategies and parameters: The SFS and SBS feature selection algorithms were tested using both “break” (stops as soon as increasing SFS or decreasing SBS results in a feature set, that has a lower recognition rate) and “best” (picks the subset consisting of the first n features – n<20 with the highest recognition rate) strategies. The kNN classifier was applied using k closest training samples in the feature space, with k varying between 3 and 8 nearest neighbors. On the other hand, the statistical LDA classifier requires no parameter input and was used with no variation. In addition to the feature selection, feature reduction based on the Fisher transformation was also applied and the recognition rates are compared to the classification results without reduction of dimensionality (Fisher vs. none).

As for the classifier evaluation methods, the normal split and random split methods were applied using both x= 0.75 and x= 0.90 parameters; In the normal split, the first x(%) of the samples are taken for training and the rest for testing, while in the random split, x(%) of the samples are taken for training and the rest for testing but the data are divided randomly. Further, in the random split method, the procedure is repeated iter times and the average recognition rate of all runs is calculated. The iter parameter was set to both 10 and 20 iterations. Finally, the one leave out method was applied with no variation, using only one sample at a time for testing and the rest to train the classifier. This is repeated for each sample and the final result is the average of all runs.

The presented choice of signal combinations, feature extraction and selection approaches as well as the classification techniques and evaluation methods adopted here are based on a pre-selection from a previous study [21]. In Zhang et al. we conducted a preliminary explicit analysis of various approaches and their combinations in order to evaluate their efficiency in the emotion recognition process [21]. For the present analysis, we only adopt the approaches which best performed in the previous study. Therefore we exclude ANOVA feature selection from the feature selection methods and the multilayer perceptron (MLP) neural network from the classification algorithms as these did not show advantageous results compared to the other approaches.

An overview of all the used physiological signals, feature selection techniques, classification algorithms, and evaluation methods is presented in Table 2.

Table 2 Overview of all used physiological signals, feature selection, classification and evaluation approaches.

Physiological Signals	- Skin Conductivity (SC) - Respiration (RSP) - SC & RSP - EMG-Corr - EMG-Zyg - EMG-Corr & EMG-Zyg - ALL signals combined
Feature Selection	- Sequential Forward Selection (SFS) - Sequential Backward Selection (SBS)
Classification Algorithms	- k-Nearest-Neighbors (kNN) - Linear Discriminant Analysis (LDA)
Evaluation Methods	- Normal Split - Random Split - One-Leave-Out

3. RESULTS

For each category-class, the classification rates result from two feature selection approaches (SFS and SBS), four classification approaches (LDA-none, LDA-Fisher, kNN-none and kNN-Fisher) combined with seven evaluation approaches (normal split 0.75, normal split 0.9, random split 0.75 10, random split 0.9 10, random split 0.75 20, random split 0.9 20 and one leave out). This is conducted for each of the seven signal configurations (SC, RSP, SC & RSP, EMG-Corr, EMG Zyg, EMG-Corr & EMG Zyg, ALL signals combined). Empty fields in the following tables were not obtained due to execution errors in the combination of the associated algorithms.

3.1. Results of the 2-category-class (Arousal)

For the 2-category-class (A: +/-), the range of classification rates varies from 36.3% to 65%. The highest recognition rate of 65% is obtained when including respiration and skin conductivity signals in the analysis. This result is obtained using the SBS feature selection, the LDA-none classification algorithm and the normal split 0.75 evaluation method. Considering all physiological signals results in a comparable recognition rate of 64.5%, using the SFS feature selection, the LDA-none classification and the normal split 0.75 evaluation method. Considering only a single physiological channel in the 2-category-class, the respiration signal seems to best contribute to the performance, with a classification rate of 64.6% obtained using the SBS feature selection, kNN-none classifier and the normal split 0.9 evaluation method. The least recognition rate of 36.3% is obtained -similar to the best recognition rate- when including respiration and skin conductivity signals in the analysis, and using the SBS feature selection and the normal split 0.75 evaluation method. However, the difference is in applying the LDA-Fisher instead of the LDA-none classification.

An overview of all classification rates obtained for the 2-category-class of Arousal is summarized in Table 3.

Table 3 Results of the 2-category-class (Arousal) classification. Best & 2nd best rates are marked in red & orange, respectively.

FeatureSelection / Classification	SFS (break)	SBS (break)	SFS (break)	SBS (break)	SFS (break)	SBS (best)	SFS (best)	SBS (break)	SFS (best)	SBS (break)	SFS (best)	SBS (best)	SFS (break)	SBS (best)
	SC		RSP		SC & RSP		EMG-Cor		EMG-Zyg		EMG-Cor & Zyg		ALL signals	
LDA-none; normal split 0.75	61%	60%	58.3%	53.7%	64.3%	65%	60.2%	54.3%	58.7%	53.3%	61.3%	53.3%	64.5%	-
LDA-none; normal split 0.9	54.2%	57.1%	54.6%	50.8%	57.1%	57.5%	58.3%	51.7%	55%	50%	53.8%	50%	52.9%	-
LDA-none; random split 0.75 10	58.1%	58.5%	50.9%	51.6%	54.9%	57.6%	51.5%	51.7%	56.3%	54.7%	53.5%	54.4%	55.1%	-
LDA-none; random split 0.9 10	59.3%	59.2%	51.9%	53.1%	56%	59.5%	52.4%	51.6%	58.1%	55.3%	54.6%	55.5%	55.6%	-
LDA-none; random split 0.75 20	57.0%	57.6%	49.4%	50.7%	53.7%	57.6%	50.2%	50.5%	55.8%	54.5%	52.8%	54.6%	54.2%	-
LDA-none; random split 0.9 20	57.9%	58.9%	49.3%	51.0%	55.9%	58.2%	52.4%	50.6%	57.3%	54.7%	54.2%	55.9%	55.5%	-
LDA-none; one leave out	56.2%	57.2%	47.8%	51.8%	54.8%	57.2%	49.8%	49.9%	55.2%	53.4%	52.4%	53.4%	53.8%	-
LDA-Fisher; normal split 0.75	39%	60%	58.3%	53.2%	59.7%	36.3%	55.8%	53.8%	41.3%	53.3%	46.2%	53.3%	62.2%	-
LDA-Fisher; normal split 0.9	54.2%	42.9%	54.6%	45.8%	51.7%	42.1%	59.2%	50%	45%	50%	50%	50%	54.2%	-
LDA-Fisher; random split 0.75 10	57.2%	55.1%	50.3%	52.6%	55.2%	56.6%	51.9%	51.9%	55.9%	54.2%	53.4%	55.3%	54.1%	-
LDA-Fisher; random split 0.9 10	56.8%	58.3%	51.7%	54.2%	56.0%	58.3%	52.5%	52.0%	55.8%	56.6%	53.9%	56.7%	55.7%	-
LDA-Fisher; random split 0.75 20	56.5%	56.1%	50.3%	51.8%	54.5%	56.3%	50.9%	50.4%	54.6%	54.1%	53.6%	54.6%	52.9%	-
LDA-Fisher; random split 0.9 20	55.7%	56.5%	49.6%	52.8%	54.8%	57.2%	52.8%	51.5%	55.5%	55.9%	53.8%	55.7%	54.4%	-
LDA-Fisher; one leave out	55.1%	53.8%	47.4%	50.7%	53.0%	55.8%	49.5%	50.0%	54.3%	53.4%	51.1%	53.4%	52.3%	-
kNN-none; normal split 0.75	58.5%	58.2%	51%	55.5%	57.2%	60.2%	47.3%	47.8%	53.3%	50.8%	49.3%	50.8%	47.8%	52.3%
kNN-none; normal split 0.9	58.3%	53.8%	55%	64.6%	60%	55.4%	47.5%	45.8%	53.8%	56.3%	50.8%	56.3%	45%	47.9%
kNN-none; random split 0.75 10	55.2%	56.7%	51.9%	52.5%	52.2%	54.8%	51.0%	49.1%	53.6%	51.5%	51.2%	51.5%	49%	52.6%
kNN-none; random split 0.9 10	57.0%	57.6%	52.8%	53.4%	53.1%	57%	51.6%	49.4%	54.0%	53.3%	54.9%	53.7%	51.0%	54.3%
kNN-none; random split 0.75 20	55.3%	55.5%	51.9%	51.0%	51.5%	54.6%	50.7%	48.6%	52.2%	50.8%	51.2%	51.7%	49.1%	52.7%
kNN-none; random split 0.9 20	56.8%	56.7%	52.8%	52.6%	52.9%	56.5%	51.7%	49%	54.2%	52.3%	53.2%	52.8%	50.3%	53.8%
kNN-none; one leave out	55.5%	55.8%	51.3%	50.6%	50.3%	54.8%	49.5%	46.3%	52.6%	51.7%	51.2%	51.7%	47.9%	51.7%
kNN-Fisher; normal split 0.75	51%	56.8%	50%	49.7%	50%	52.8%	46.5%	50.8%	50%	50.8%	52.8%	50.8%	56.3%	-
kNN-Fisher; normal split 0.9	55.4%	48.8%	57.1%	55.4%	56.3%	50%	55.4%	50%	59.6%	56.3%	49.6%	56.3%	58.3%	-
kNN-Fisher; random split 0.75 10	54.9%	54.8%	51.9%	51.2%	52.0%	54%	51.5%	51.1%	52.2%	51.4%	52.7%	51.5%	52.2%	-
kNN-Fisher; random split 0.9 10	54.8%	55.1%	53.0%	52.4%	56.3%	54.5%	51.2%	52.1%	54.1%	52.5%	52.6%	54.1%	55.0%	-
kNN-Fisher; random split 0.75 20	54.2%	54.2%	50.7%	50.5%	52.3%	53.6%	49.9%	51.1%	51.2%	51.0%	52.3%	51.0%	51.6%	-
kNN-Fisher; random split 0.9 20	54.9%	53.9%	51.7%	51.4%	53.3%	54.7%	52.9%	51.7%	52.9%	52.2%	52.8%	53.0%	52.6%	-
kNN-Fisher; one leave out	59.6%	53.9%	50%	50.3%	52.9%	49.7%	53.1%	51.3%	51.1%	51.7%	49.3%	51.7%	50.9%	-

3.2. Results of the 3-category-class (Valence)

For the 3-category-class defined by the valence dimension (V: +/-/0), the classification rates range from 14.8% to 45.7%. The performance is always better when fusing two or more physiological signals than considering single signal channels. The highest recognition rate of 45.7% appears four times and seems to give a robust predication about the best performing methods for this category class for the differentiation of various valence states. This result is obtained when using both LDA-none and LDA-Fisher classification algorithms each combined with the SFS feature selection approach and the normal split 0.75 evaluation method. Further, this result is obtained for the skin conductivity signal combined with the respiration signal channels as well as when fusing all physiological signals together. The least recognition rate of 14.8% is obtained when only considering the skin conductivity signal channel and applying the LDA-Fisher classification algorithm combined with the SFS feature selection approach and the one-leave-out evaluation method.

An overview of all classification rates obtained for the 3-category-class of Valence is summarized in Table 4.

Table 4 Results of the 3-category-class (Valence) classification. Best & 2nd best rates are marked in red & orange, respectively.

FeatureSelection	SFS (break)	SBS (break)	SFS (break)	SBS (best)	SFS (break)	SBS (best)	SFS (break)	SBS (break)	SFS (break)	SBS (best)	SFS (break)	SBS (best)	SFS (break)	SBS (best)
Classification	SC		RSP		SC & RSP		EMG-Cor		EMG-Zyg		EMG-Cor & Zyg		ALL signals	
LDA-none; normal split 0.75	40.7%	39%	42.3%	43%	45.7%	38.3%	44.3%	39.3%	40%	33.3%	44%	33.3%	45.7%	33.3%
LDA-none; normal split 0.9	42.5%	43.3%	36.7%	40.8%	42.5%	41.7%	43.3%	41.7%	39.2%	33.3%	28.3%	33.3%	32.5%	33.3%
LDA-none; random split 0.75 10	38.4%	38.2%	33.5%	37.1%	34.7%	35%	38.1%	36.7%	39.6%	33.3%	36.8%	33.3%	37.3%	33.3%
LDA-none; random split 0.9 10	39.9%	40.2%	35.7%	38.2%	37.7%	36.2%	39.7%	36.8%	41.4%	33.3%	39.1%	33.3%	39.8%	33.3%
LDA-none; random split 0.75 20	38.5%	38.6%	33.4%	36.8%	34.6%	34.3%	37.9%	35.7%	38.7%	33.3%	36.4%	33.3%	36%	33.3%
LDA-none; random split 0.9 20	39.3%	39.3%	33.5%	39.1%	36.1%	35.1%	39.6%	36.1%	41.4%	33.3%	38%	33.3%	37.8%	33.3%
LDA-none; one leave out	37.5%	37.8%	29.5%	36.3%	33%	35.5%	37.9%	35.1%	38.4%	33.3%	36.3%	33.3%	36.3%	33.3%
LDA-Fisher; normal split 0.75	40.7%	35.7%	37%	43%	45.7%	40.7%	44.3%	34%	40%	-	44%	-	45.7%	-
LDA-Fisher; normal split 0.9	32.5%	41.7%	35.8%	40.8%	42.5%	40%	43.3%	39.2%	39.2%	-	28.3%	-	32.5%	-
LDA-Fisher; random split 0.75 10	37.6%	39.4%	34.0%	37.5%	35.1%	37.2%	38.1%	34.8%	39.8%	-	36.7%	-	37.2%	-
LDA-Fisher; random split 0.9 10	37.7%	41.1%	34.9%	38.3%	37.1%	38.9%	38.8%	37.3%	42.2%	-	37.9%	-	38%	-
LDA-Fisher; random split 0.75 20	37.3%	37.9%	33.6%	37.2%	35.0%	37.3%	37.2%	35.9%	38.8%	-	36.4%	-	36.5%	-
LDA-Fisher; random split 0.9 20	36.5%	39.8%	35.0%	38.2%	36.5%	38.4%	39.0%	37.1%	40.7%	-	37%	-	37.0%	-
LDA-Fisher; one leave out	14.8%	37.4%	32.5%	36.3%	33%	35.4%	38.4%	35.8%	38.4%	-	36.3%	-	36.3%	-
kNN-none; normal split 0.75	36.7%	38%	37.3%	32.7%	40.3%	34.7%	33.3%	33.7%	37.3%	34.7%	39%	34.7%	38.3%	34.7%
kNN-none; normal split 0.9	39.2%	37.5%	40%	33.3%	39.2%	40.8%	33.3%	32.5%	37.5%	33.3%	37.5%	35%	35.8%	35%
kNN-none; random split 0.75 10	37.8%	38.3%	34%	35.4%	35.7%	36.3%	33.8%	35.3%	37.1%	35.3%	37.0%	35.5%	35.8%	36.1%
kNN-none; random split 0.9 10	40.1%	39.6%	36.9%	37.4%	36.7%	37.6%	34.1%	37.1%	38.5%	36.7%	37.7%	37.6%	38.5%	35.5%
kNN-none; random split 0.75 20	37.3%	37%	34.6%	35.3%	35.0%	34.8%	33.5%	34.3%	36.4%	34.1%	36.8%	34.0%	36.1%	34.6%
kNN-none; random split 0.9 20	40.1%	37.8%	34.8%	37.4%	36.3%	36.3%	33.9%	36.2%	37.8%	36.5%	37.7%	35.7%	37%	35.7%
kNN-none; one leave out	35.2%	35.3%	34.8%	36%	34.3%	33.6%	32.5%	34.1%	37.3%	34.8%	36.1%	34.8%	35%	34.8%
kNN-Fisher; normal split 0.75	36.7%	39.3%	40%	36.3%	36.3%	41%	35%	36.7%	37%	-	38.7%	-	35%	-
kNN-Fisher; normal split 0.9	35.8%	42.5%	35%	43.3%	37.5%	40.8%	32.5%	40.8%	30.8%	-	35%	-	36.7%	-
kNN-Fisher; random split 0.75 10	37.7%	39.5%	34.3%	36.4%	36%	38%	35.3%	36.0%	37.3%	-	36.3%	-	35.6%	-
kNN-Fisher; random split 0.9 10	38.5%	40.5%	34.1%	37.9%	37.2%	38.9%	37.8%	37.9%	37.6%	-	37.8%	-	36.4%	-
kNN-Fisher; random split 0.75 20	36.9%	38.9%	34.1%	34.7%	35.0%	36.1%	35.0%	35.6%	35.6%	-	35.3%	-	35.0%	-
kNN-Fisher; random split 0.9 20	37.8%	40.0%	35.8%	36.7%	36.1%	38.4%	35.3%	36.6%	36.1%	-	36.7%	-	35.2%	-
kNN-Fisher; one leave out	36.3%	38.8%	33.6%	34.9%	34.8%	36.8%	33.6%	36%	35.5%	-	35.3%	-	33.7%	-

3.3. Results of the 5-category-class (Valence/Arousal)

For the 5-category-class defined by the valence and arousal dimensions (VA: 0-/++/-+/+-/--), the range of classification rates varies from 16% to 36%. We obtained the highest recognition rates of 36% when combining the skin conductivity signal with the respiration signal. This is obtained using the SBS feature selection approach with the kNN-Fisher classification algorithm and the normal split 0.9 evaluation method. Fusing all physiological signals together, resulted in a recognition rate of 32.4%. This was obtained using the SFS feature selection approach and the LDA-none classification algorithm combined with the normal split 0.75 evaluation method. When considering only single channels, the performance is always below 30%, which is below the rates obtained when fusing two or more signal channels. The least recognition rate of 16% is obtained when only considering the EMG-Corrugator muscle signal and applying the LDA-Fisher classification algorithm combined with the SFS feature selection approach and the normal split 0.9 evaluation method.

An overview of all classification rates obtained for the 5-category-class of Valence and Arousal is summarized in Table 5.

Table 5 Results of the 5-category-class classification. Best & 2nd best rates are marked in red & orange, respectively.

Feature Selection / Classification	SFS (best)	SBS (break)	SFS (break)	SBS (best)	SFS (break)	SBS (best)	SFS (break)	SBS (break)	SFS (best)	SBS (best)	SFS (best)	SBS (best)	SFS (break)	SBS (best)
	SC		RSP		SC & RSP		EMG-Cor		EMG-Zyg		EMG-Cor & Zyg		ALL signals	
LDA-none; normal split 0.75	28%	27.6%	27.6%	28.8%	32.4%	35.6%	25.6%	26.4%	26.4%	20%	30.4%	20%	32.4%	20%
LDA-none; normal split 0.9	27%	29%	22%	20%	27%	30%	20%	26%	22%	20%	26%	20%	27%	20%
LDA-none; random split 0.75 10	23.8%	23.0%	19.5%	22%	22.9%	24.2%	21.3%	23.0%	25.1%	20%	22.5%	20%	22.8%	20%
LDA-none; random split 0.9 10	25.2%	22.8%	21.3%	23%	24%	25.6%	21.4%	23.4%	26%	20%	23%	20%	24.3%	20%
LDA-none; random split 0.75 20	23.7%	22.3%	19.9%	21.3%	21.8%	24.3%	21.5%	22.5%	24.4%	20%	22.1%	20%	22.2%	20%
LDA-none; random split 0.9 20	24.6%	23.1%	20.4%	23.1%	22.4%	25.4%	21.4%	23.2%	25.9%	20%	23.9%	20%	23.0%	20%
LDA-none; one leave out	24.2%	22.7%	19.7%	21.5%	21.5%	24.6%	19%	22.1%	23.4%	20%	21.6%	20%	21.5%	20%
LDA-Fisher; normal split 0.75	25.2%	24.8%	23.2%	26.4%	26%	29.6%	23.2%	22%	24%	-	23.2%	-	26%	-
LDA-Fisher; normal split 0.9	26%	25%	20%	25%	26%	30%	16%	29%	19%	-	25%	-	26%	-
LDA-Fisher; random split 0.75 10	24.8%	23.6%	21.4%	21.9%	23.5%	25.3%	21.9%	22.3%	23.7%	-	22.5%	-	24.1%	-
LDA-Fisher; random split 0.9 10	26.5%	23.9%	21.5%	24.3%	25.7%	27.2%	22.9%	22.8%	24.4%	-	23.2%	-	24.6%	-
LDA-Fisher; random split 0.75 20	23.8%	23.1%	19.9%	21.3%	22.9%	24.4%	21.3%	21.6%	22.9%	-	21.0%	-	23.2%	-
LDA-Fisher; random split 0.9 20	25.4%	23.9%	20.8%	23%	24.1%	26.1%	22.8%	22.8%	23.6%	-	21.7%	-	24.2%	-
LDA-Fisher; one leave out	24.7%	23.6%	20.7%	21.6%	22.7%	24.5%	21.7%	19%	17.6%	-	20.1%	-	22.7%	-
kNN-none; normal split 0.75	22%	24.4%	19.2%	24%	28%	22.4%	22.4%	22.4%	23.2%	18.8%	22.4%	18.8%	28%	18.8%
kNN-none; normal split 0.9	25%	25%	19%	29%	28%	18%	21%	26%	24%	21%	24%	21%	28%	21%
kNN-none; random split 0.75 10	24.6%	23.8%	18.8%	23.4%	22.9%	23.5%	23.5%	22.2%	22.0%	21.0%	22.6%	21.0%	22.7%	21.5%
kNN-none; random split 0.9 10	26.1%	25%	19.8%	25.1%	24.3%	24%	23.7%	23.1%	22.4%	21.8%	25.2%	23.2%	24.2%	22%
kNN-none; random split 0.75 20	24.8%	23.8%	17.9%	23.4%	22.6%	21.8%	23.1%	20.9%	21.4%	20.4%	22.3%	20.4%	22.9%	20.8%
kNN-none; random split 0.9 20	25.6%	24.2%	18.7%	23.7%	23.8%	22.9%	24.2%	22.5%	22.3%	21.3%	23.1%	21.3%	23.9%	21.4%
kNN-none; one leave out	23.9%	22.9%	16.9%	22.4%	21.8%	21.4%	22.2%	20.6%	20.7%	19.5%	22.5%	19.5%	21.8%	19.5%
kNN-Fisher; normal split 0.75	24.4%	22%	21.6%	24.4%	24.8%	23.2%	24.4%	24.4%	23.6%	-	25.2%	-	24.8%	-
kNN-Fisher; normal split 0.9	23%	21%	23%	24%	27%	36%	22%	24%	24%	-	21%	-	27%	-
kNN-Fisher; random split 0.75 10	23.6%	23.0%	21.9%	22%	23.0%	23.0%	22.7%	22.5%	22.6%	-	22.2%	-	22.8%	-
kNN-Fisher; random split 0.9 10	25%	23.8%	22.9%	23.6%	23.6%	25.4%	23.3%	23.8%	24.7%	-	23%	-	24.5%	-
kNN-Fisher; random split 0.75 20	23.2%	23.3%	20.8%	21.5%	21.7%	23.0%	21.5%	21.8%	22.4%	-	21.0%	-	22.0%	-
kNN-Fisher; random split 0.9 20	24.9%	23.4%	22.0%	22.1%	22.1%	24.2%	22.5%	23%	24.8%	-	21.7%	-	23.0%	-
kNN-Fisher; one leave out	24.1%	22%	22.2%	20.3%	21.5%	23%	20.7%	22.9%	22.6%	-	20.3%	-	21.5%	-

4. DISCUSSION

Overall, the classification results show recognition rates of up to 65%, 45.7% and 36% obtained for the 2-category-class, 3-category-class and 5-category-class, respectively. For all the category-classes, the best results are obtained when considering skin conductivity combined with the respiration signals. However, including all the signals resulted in very comparable highest rates of 64.5%, 45.7% and 32.4% for the 2-category-, 3-category- and 5-category-classes, respectively. All three best rates for ALL signals combined are obtained using the SFS feature selection, the LDA classifier and the normal split evaluation method, showing a robust combination of approaches leading to the best performance. Table 6 summarizes the combination of approaches resulting in the best recognition rates and thus presenting the highest performance in terms of classification rates. The results are illustrated for all the three different category-classes and for all signal configurations defined and investigated in this study.

Although the best recognition results are above the probability for random hits due to chance (50%, 33.3% and 20%, respectively) the highest rates present only a satisfactory result. Also large deviations within the results of each of the category-classes could be observed. The resulting least classification rates of all category-classes are all obtained with the LDA-Fisher classification, that is, when the feature reduction technique using the Fisher transformation is applied. Especially for the 2-category-class of Arousal, the least recognition rate of 36.3% is observed using the SBS feature selection, the LDA-Fisher classification and the normal split 0.75 evaluation method, for the skin conductivity combined with the respiration signals. This finding is interesting because the highest classification rate for the 2-category-class is obtained using exactly the same combination of approaches, but without the Fisher transformation. This result reminds us that high caution must be taken when reducing the dimension of the original data in order to decrease the execution time and space complexity. While transforming the high dimensional data into low dimensional data, some important information will be lost. The results we obtained show that the Fisher transformation is not suitable in all cases and that other feature reduction techniques (e.g. Principal Component Analysis (PCA)) should be further investigated on their performance.

Further, the respiration channel seems to best contribute to the classification rates of the 2-category-class. This can be explained by the fact that respiration is quite sensitive to the level of arousal, which means people breathe fast and deep when they are highly aroused. On the other side, the EMG-zygomaticus signal seems to make the least contribution on the classification performance compared to the other physiological signals. Compared to the single channels, multi-channels recognition, combining two or more signals, shows advantageous classification performance. This finding is reasonable and consistent with current researchers' results.

Table 6. Approaches with the highest performance in terms of classification rates.

Cat \ Signals	SC	RSP	SC & RSP	EMD-Cor	EMG-Zyg	EMG-Cor & Zyg	ALL signals
2 Categories	**61%**	**64.6%**	**65%**	**60.2%**	**59.6%**	**61.3%**	**64.5%**
	SFS	SBS	SBS	SFS	SFS	SFS	SFS
	LDA-none	kNN-none	LDA-none	LDA-none	kNN-Fisher	LDA-none	LDA-none
	normal split 0.75	normal split 0.9	normal split 0.75	normal split 0.75	normal split 0.9	normal split 0.75	normal split 0.75
3 Categories	**43.3%**	**43.3%**	**45.7%**	**44.3%**	**42.2%**	**44%**	**45.7%**
	SBS	SBS	SFS	SFS	SFS	SFS	SFS
	LDA-none	kNN-Fisher	LDA-none/Fisher	LDA-none/Fisher	LDA-Fisher	LDA-none/Fisher	LDA-none/Fisher
	normal split 0.9	normal split 0.9	normal split 0.75	normal split 0.75	random split 0.9	normal split 0.75	normal split 0.75
5 Categories	**29%**	**29%**	**36%**	**29%**	**26.4%**	**30.4%**	**32.4%**
	SBS	SBS	SBS	SBS	SFS	SFS	SFS
	LDA-none	kNN-none	kNN-Fisher	LDA-Fisher	LDA-none	LDA-none	LDA-none
	normal split 0.9	normal split 0.9	normal split 0.9	normal split 0.9	normal split 0.75	normal split 0.75	normal split 0.75

5. Conclusion

In this study, we present a classification analysis using various combinations of approaches and physiological signals for different emotional category classes. The goal was to evaluate the performance of those approaches in terms of classification rates for the recognition of human affective states. Highest rates obtained when fusing all physiological signals resulted when applying the SFS feature selection, LDA classifier and normal split evaluation method, showing a robust combination of approaches leading to the best performance. On the other hand, the Fisher transformation seems to be not always advantageous for the performance rates. In future, further analysis could include other classification algorithms, such as support vector machine, random forest or Bayesian classifiers to obtain higher and more stable results. Also an extensive feature analysis could also be performed to investigate the effects of individual features on the classification performance.

Acknowledgements

The authors´ research was supported by the Transregional Collaborative Research Center SFB/TRR 62 Companion Technology for Cognitive Technical Systems funded by the German Research Foundation (DFG). It is also supported by a Margarete von Wrangell (MvW) Habilitation scholarship funded by the Ministry of Science, Research and Arts (MWK) of the state of Baden-Württemberg for Dr. Dilana Hazer-Rau and a doctoral scholarship funded by the China Scholarship Council (CSC) for Lin Zhang.

REFERENCES

[1] H. Hoffmann, H.C. Traue, F. Bachmayr, & H. Kessler, “Perceived realism of dynamic facial expressions of emotion: optimal durations for the presentation of emotional onsets and offsets”, Cognition & Emotion, 24(8), 1369-1376, 2010.

[2] R.A. Calvo, S. D’Mello, “Affect Detection: An interdisciplinary Review of Models, Methods, and Their Applications.” IEEE Transactions on affective computing, 1(1), 2010.

[3] P. Patel, A. Chaudhari, R. Kale, M.A.Pund, "Emotion recognition from speech with gaussian mixture models & via boosted GMM" International Journal of Research In Science & Engineering IJRISE, Vol. 3, Issue 2, 2017.

[4] A. Pentland, & S. Pentland, Honest signals: how they shape our world. Bradford Books. MIT Press, 2008.

[5] R.W. Picard, “Affective computing: challenges.” International Journal of Human-Computer Studies, 59(1), 55-64, 2003.

[6] S. Walter, J. Kim, D. Hrabal, S.C. Crawcour, H. Kessler, & H.C. Traue, “Transsituational individual-specific biopsychological classification of emotions.” Systems, Man, and Cybernetics: Systems, IEEE Transactions on, 43(4), 988-995, 2013.

[7] Y.L. Hsu, J.S. Wang, W.C. Wang, C.H. Hung, "Automatic ECG-Based Emotion Recognition in Music Listening." IEEE Transactions on Affective Computing, 2017.

[8] J.A. Healey and R.W. Picard, “Detecting stress during realworld driving tasks using physiological sensors,” IEEE Transactions on Intelligent Transportation Systems, vol. 6, p. 156166, 2005.

[9] W. Wei, Q. Jia, Y. Feng, G. Chen, “ Emotion Recognition Based on Weighted Fusion Strategy of Multichannel Physiological Signals." Computational Intelligence and Neuroscience, Article ID: 5296523, 2018.

[10] J. Kim, “Bimodal emotion recognition using speech and physiological changes,” in Robust Speech Recognition and Understanding, M. Grimm and K. Kroschel, Eds. I-Tech Education and Publishing, Vienna, Austria, pp. 265–280, 2007.

[11] C. Frantzidis, C. Bratsas, C. Papadelis, E. Konstandinidis, C. Pappas, and P. Bamidis, “Towards emotion aware computing: An integrated approach using multi-channel neuro-physiological recordings and affective visual stimuli,” IEEE Trans. on Information Technology in Biomedicine, vol. 14, no. 3, pp. 589, 2010.

[12] G. Chanel, J. J. Kierkels, M. Soleymani, and T. Pun, “Short-term emotion assessment in a recall paradigm,” International Journal of Human-Computer Studies, vol. 67, p. 607627, 2009.

[13] M. Soleymani, G. Chanel, L. J. M. Kierskels, and T. Pun, “Affective characterization of movie scenes based on content analysis and physiological changes,” International Journal of Semantic Computing, vol. 3, pp. 235–254, 2009.

[14] J.J. McGinley, B.H. Friedman, "Autonomic specificity in emotion: The induction method matters." International Journal of Psychophysiology, vol. 118: pp. 48-57, 2017.

[15] M. Uhrig, N. Trautmann, U. Baumgärtner, R.D. Treede, F. Henrich, W. Hiller & S. Marschall, "Emotion elicitation: A comparison of pictures and films." Frontiers in psychology, vol. 7, 2016.

[16] S. Walter, H. Kessler, S. Gruss, L. Jerg-Bretzke, A. Scheck, J. Ströbel, H. Hoffmann & H.C. Traue, “The influence of neuroticism and psychological symptoms on the assessment of images in three-dimensional emotion space”. GMS Psychosocial Medicine, vol. 8: Doc04, 2011a.

[17] P. Lang, M. Bradley & B. Cuthbert, “International affective picture system (iaps): Affective ratings of pictures and instruction manual,” University of Florida, Gainesville, FL, USA, Tech. Rep., 2008.

[18] G. Valenza, A. Lanata, and E.P. Scilingo, "The Role of Nonlinear Dynamics in Affective Valence and Arousal Recognition". IEEE Transactions on Affective Computing, PrePrints. 99, 2011.

[19] J. Wagner, Augsburg biosignal toolbox (Aubt) – User Guide. University of Augsburg, 2006.

[20] J. Wagner, J. Kim, E. André, "From Physiological Signals to Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification." In IEEE International Conference on Multimedia & Expo (ICME), 2005.

[21] L. Zhang, S. Rukavina, S. Gruss, H.C. Traue, D. Hazer, "Classification analysis for the emotion recognition from psychobiological data." In: International Symposium on Companion-Technology, ISCT, IEEE, 2015.